Zero-shot classification of ECG signals using a CLIP-based model

Navmeet S. Jassal, Konstantin Egorov\*, Semen BUDENNYY,2

1Laboratory X, Institute X, Department X, Organization X, City X, State XX (only USA, Canada and Australia), Country

2Laboratory X, Institute X, Department X, Organization X, City X, State XX (only USA, Canada and Australia), Country

**\* Correspondence:**Corresponding Author  
ndzhassal2@edu.hse.ru

Keywords: keyword1, keyword2, keyword3, keyword4, keyword5. (Min.5-Max. 8)

Abstract

For full guidelines please refer to [Author Guidelines](https://www.frontiersin.org/guidelines/author-guidelines)

As a primary goal, the abstract should render the general significance and conceptual advance of the work clearly accessible to a broad readership. References should not be cited in the abstract. Leave the Abstract empty if your article does not require one – please see the “Article types” on every Frontiers journal page for full details.

# Introduction

According to the World Heart Federation (WHF), prominent medical institutions, and leading healthcare professionals, cardiovascular diseases (CVDs) are a leading cause of mortality worldwide [17; 22; 25; 36]. In the United States, the Center for Disease Control (CDC) estimates that the annual total cost associated with CVD-related treatment and mortality is approximately $219 billion dollars (USD) per year [18; 21]. Given the significant financial and socioeconomic burdens of CVDs, it is essential that individuals receive timely and accurate medical care for the diagnosis, treatment, and ongoing care of CVDs.

The electrocardiogram (ECG) remains the medical standard and benchmark for the identification and diagnosis of CVDs with more than 300 million ECGs being performed globally [7]. An ECG measures the changes in electrical membrane potential of the heart across different directions of the body and are often referred to as leads with a 12-lead ECG report being the most common type of ECG clinical report [14; 29]. Once an ECG report is available, it is then analyzed by a licensed medical professional such as a cardiologist or an auxiliary medical professional.

Interpreting and analyzing ECG reports remains a highly complex and time-consuming process for medical professionals [11]. This process not only requires manual annotation of ECGs but also mandates that the individual analyzing the ECG possesses a vast understanding of topics such as cardiac anatomy, electrophysiology, pattern recognition, coronary distribution, and pathophysiology to perform correctly and accurately [11]. A systematic literature review by Cook et al. revealed that medical professionals, such as cardiologists, across levels of education experienced challenges in providing a correct diagnosis for an ECG report with correct interpretation accuracy ranging from 49% to 92% and with a median interpretation accuracy of 57% [11]. The review also found that medical professionals who received continuing education and training improved their ability to correctly provide a diagnosis for an ECG report with a median accuracy of 67%, underscoring the importance of continuous professional development in ECG interpretation [11].

Recognizing the challenges in ECG interpretation faced by medical professionals, researchers have turned to advancements in machine learning (ML), deep learning (DL), and artificial intelligence (AI) to address this issue. Namely, with the numerous advancements observed in ML, DL, and AI vision models over recent years, it was postulated that ML and DL vision models could accurately classify pathologies from ECG reports.

In this paper, we aim to advance the field of deep learning (DL) as it pertains to ECG classification. Specifically, we demonstrate the effectiveness of using a multimodal Contrastive Language-Image Pre-training (CLIP) framework for diagnosing various combinations and permutations of ECG diagnostic classes. Our literature review indicates that the application of a multimodal CLIP framework in ECG classification has been minimally explored, with only a few studies published in the past two years.

In the next section, “Related Works,” we provide an overview of CLIP and the current state of ML and DL models for ECG classification. Additionally, we discuss Zero-shot classification as it relates to our study.

# Related Works

## Contrastive language image pre-training (CLIP)

In 2021, OpenAI researchers introduced a novel approach to multimodal learning in their paper “Learning Transferable Visual Models from Natural Language Supervision,” which presented the Contrastive Language-Image Pre-training or CLIP for short [10; 20]. The CLIP framework consists of two components: an image encoder (vision model) and a text encoder (text model) [10; 20]. Within a CLIP framework, both encoders are trained together, allowing the CLIP-based model to develop a deep understanding of its image-text pairs [10]. To enable the models to learn, the CLIP framework employs a function called contrastive loss (also known as symmetric loss) [10; 20]. This loss function calculates the distance between the embeddings of image-text pairs, minimizing the distance for similar pairs (positive instances) and maximizing it for dissimilar pairs (negative instances) [10; 20]. In doing this, the image and text models are incentivized to learn better representations of their inputs to capture the semantic relationship between embeddings of image-text pairs, while also penalizing or disincentivizing poor representations of the image-text pair inputs [20].

Unlike traditional ML or DL approaches for ECG classification, which typically involve training an image feature extractor alongside a classifier to predict image classes or labels, the CLIP-based model offers a novel approach for ECG classification tasks [20]. This is particularly interesting and useful for ECG classification because a CLIP-based model can be utilized for zero-shot classification, allowing it to handle a variety of classification tasks and datasets without the need for task-specific training. This is something that current state-of-the-art (SOTA) ML or DL approaches in the domain of ECG classification cannot do.

## Related Works – Deep Learning and ECG Classification

In the fields of AI, ML, and DL, a primary objective has been to develop models that can accurately classify cardiovascular diseases (CVDs) from ECG reports and/or ECG signals [2]. To achieve this, a patient’s medical history, including their ECG reports, is aggregated, preprocessed, and analyzed by the model, which then provides a recommendation or classification based on the input data.

In the past decade, several notable ECG classification models have been developed that have demonstrated to be as effective in correctly classifying CVDs from ECGs as experienced cardiologists using ECG signal inputs [4; 23]. A systematic literature review by Petmezas et al. found that the most popular deep learning models for ECG classification include variants of convolutional neural networks (CNNs), recurrent neural networks (RNNs), ResNet models, and long short-term memory (LSTM) models [33]. Among the studies analyzed in this review, 66.1% of the studies favored using CNNs for specific ECG classification tasks [33]. For instance, Lu et al. achieved an accuracy of 99.31% for arrhythmia classification using a 1D-CNN [19; 33]. Similarly, Yu et al. reported an accuracy of 99.70% for premature ventricular contraction using a 1D-CNN [6; 33]. Both examples exemplify the robustness of CNNs for various ECG classification tasks and many other studies analyzed in this review also demonstrated that CNNs were effective for other classes such as myocardial infarction, coronary artery disease, and congestive heart failure [33].

## Related Works – CLIP-based models in medical imaging

CLIP-based models have proven effective across a wide range of datasets, including Food-101, CIFAR-100, ImageNet, and MNIST [20]. However, their utilization in medical imaging for image classification or diagnostic tasks remains an area of research that has not been extensively explored [9]. In their systematic literature review on the topic of CLIP-based models for medical imaging, Zhao et al. found that since 2021, there have been approximately 38 studies that have utilized the CLIP-based model framework for medical imaging tasks [9]. Of the 38 studies presented in this review, notable examples include seven implementations focusing on chest X-ray classification, one on stomach histology, one on lung CT scans, one on skin/dermatology conditions, one on brain MRIs, and one on eye classification tasks [9]. This emerging field shows promise, but further research is needed to fully understand and leverage the potential of CLIP-based models in medical imaging, in particular, for ECG classification.

In Zhao et al.’s systematic review, only one study utilized the CLIP-based framework for ECG classification tasks, using ECG signals as the primary input [9; 15]. In their paper, “ETP: Learning Transferable ECG Representations via ECG-Text Pre-Training,” Liu et al. describe training a CLIP-based model from scratch using the PTB-XL and CPSC2018 datasets [3; 15; 30]. In this study, authors employed 12-lead ECG reports and their corresponding signals to train a model where a 1D-ResNet18 served as the image encoder and BioClinicalBERT was used as the text encoder [15; 31]. During training, the 1D-ResNet18 model’s weights were updated, while the BioClinicalBERT’s weights remained frozen [15]. The aim here was to demonstrate the viability of zero-shot ECG classification tasks [15].

Liu et al.'s experiments achieved state-of-the-art (SOTA) results for ECG classification on the PTB-XL and CPSC2018 datasets during both supervised learning and zero-shot classification [15]. In the PTB-XL zero-shot classification experiment, their ETP model achieved higher AUC, ACC, and F1 scores across four distinct classes of ECGs [15]. Similarly, the ETP approach yielded significant results on the CPSC2018 dataset, achieving higher AUC, ACC, and F1 scores across nine distinct classes of ECGs [15]. Overall, both experiments and their corresponding findings demonstrated the robustness of the CLIP-based framework for ECG classification in both supervised and zero-shot learning scenarios [15]. This study is significant as it was the first to demonstrate the feasibility of using the CLIP framework for zero-shot classification of ECGs. Building on Liu et al.'s work, our study aims to extend this approach across a greater number and by extension combinations of ECG diagnostic classes.

Our literature search also identified another study where an implementation of the CLIP framework was presented for an ECG classification task. In Li et al.’s “Frozen Language Model Helps ECG Zero-Shot Learning”, the authors introduced a new technique called Multimodal ECG-text SSL or METS for short [16]. Like Liu et al.'s approach, Li et al. used a text encoder with frozen weights where the goal of METS was to leverage automatically generated ECG clinical reports to guide the training of the ECG image encoder [15; 16]. The METS approach demonstrated a 10% performance improvement without the use of annotated datasets, compared to other studies in the field [16]. Overall, the METS model processes an ECG signal and its corresponding clinical report/diagnosis text within a multimodal learning framework, akin to CLIP [16; 20]. Using contrastive loss, the model measured the distance between image-text pairs, enabling the ECG image encoder to learn “deep” representations of its inputs [16]. When evaluated, the METS approach achieved higher accuracy, precision, recall, and F1 scores on the PTB-XL dataset compared to other leading methods [16].

## Related Works – Zero-shot Learning & Classification

In ML and DL research, zero-shot classification refers to the capability of a trained model to recognize and correctly classify new, unseen objects or information based on its input [39, 40]. One benefit of zero-shot learning is when working with many classes and even subclasses of labeled data. In production level datasets or databases, it is unlikely that every single datapoint will have an associated label or class. As such, training a ML or DL model in a zero-shot learning condition can be valuable in these cases as it may be able to recognize the unlabeled and/or previously unseen data.

The zero-shot learning framework involves two main steps: model training and model inference [39]. During the training phase, the model is trained on a set of known classes from a dataset, where the model is expected to learn the representation of its inputs [39]. In the model inference phase, the trained model is then evaluated on classes of data that were not included during the training phase, testing the model’s ability to generalize to new, unseen classes of data [39; 40].

In our study, we employ a CLIP-based model within a zero-shot learning framework. It is hypothesized that a CLIP-based model is well-suited for zero-shot classification for two reasons. First, a CLIP-based model consisting of an image and text encoder are jointly trained on ECG signals and their classes, respectively. This allows the CLIP-based model to learn deep, semantic representations of the image-text pair inputs. Secondly, in the CLIP-based model, we utilize the contrastive loss function which enables the model to effectively group image-text pair embeddings which are similar closer together and vice versa. In our study, the CLIP-based model will be trained on a specific set of ECG diagnostic classes. Then, during the evaluation phase, the trained CLIP-based model will be tested on ECG diagnostic classes that were excluded during the training phase. This zero-shot learning framework allows us to assess the CLIP-based model’s ability to generalize to new, unseen classes of ECGs.

# Materials and Methodology

## Datasets

Two experiments were performed in this study, Experiments A and B, respectively.

### Datasets – Experiment A

In Experiment A, a CLIP-based model consisting of the BioClinicalBERT text encoder and a 1D-CNN image encoder, is trained on the Physikalisch-Technische Bundesanstalt (PTB-XL) dataset. PTB-XL is a large publicly available electrocardiography dataset containing 21,799 records \citealp{physionetdataset1}. Each record contains multiple types of data, however, the three main forms of data utilized in Experiment A are: the raw 10 second 12-lead ECG signal, the associated ECG clinician's report, and the class label for the associated ECG \citealp{physionetdataset1}. The goal of Experiment A is determine the efficacy and viability of employing a CLIP-based model within the framework of zero-shot learning to evaluate the model's performance in classifying ECG diagnostic classes that were excluded during the training of the CLIP-based model.

During training, the CLIP-based model is trained on the waveform ECG signals and their corresponding ECG clinician's reports which serve as the inputs for the 1D-CNN and BioClinicalBERT, respectively. After the CLIP-based is trained, it will be evaluated on the following set of excluded classes: Left ventricular hypertrophy, ST depression, Low QRS voltages, ST changes, T wave abnormal, and Right axis deviation. A full breakdown of all ECG diagnostic class labels and their total samples can be found in the Supplementary Materials section.

To evaluate the model's performance during both training and inference phases, the evaluation metrics utilized in Experiment A are accuracy score, Area under the ROC Curve (AUC-ROC), and Area under the precision-recall curve (AUC-PR). In the PTB-XL dataset, ECG signals can contain multiple diagnostic class labels. As such, to calculate accuracy score, we employ the following condition: if the model is able to correctly classify at least one label for a ECG image-text pair containing multiple diagnostic classes, then it is considered a correct classification. If it does not correctly classify at least one label, then it is considered an incorrect classification.

**Figure 1**: Distribution of Classes (labels) in the PTB-XL dataset

|  |  |
| --- | --- |
| **Diagnostic Class** | **Total Samples** |
| sinus rhythm | 18058 |
| myocardial infarction | 5248 |
| left axis deviation | 5146 |
| abnormal QRS | 3389 |
| left ventricular hypertrophy | 2354 |
| t wave abnormal | 2341 |
| myocardial ischemia | 2171 |
| left anterior fascicular block | 1624 |
| atrial fibrillation | 1514 |
| ventricular ectopics | 1153 |
| incomplete right bundle branch block | 1118 |
| st depression | 1009 |
| sinus tachycardia | 826 |
| 1st degree av block | 795 |
| nonspecific intraventricular conduction disorder | 787 |
| sinus arrhythmia | 772 |
| s t changes | 768 |
| sinus bradycardia | 637 |
| qwave abnormal | 548 |
| complete right bundle branch block | 542 |
| left bundle branch block | 536 |
| left atrial enlargement | 426 |
| premature atrial contraction | 398 |
| nonspecific st t abnormality | 381 |
| anterior myocardial infarction | 353 |
| right axis deviation | 343 |
| prolonged pr interval | 340 |
| t wave inversion | 294 |
| pacing rhythm | 294 |
| inferior ischaemia | 218 |
| low qrs voltages | 182 |
| left posterior fascicular block | 177 |
| supraventricular premature beats | 157 |
| indeterminate cardiac axis | 155 |
| lateral ischaemia | 140 |
| right ventricular hypertrophy | 126 |
| prolonged qt interval | 118 |
| right atrial hypertrophy | 99 |
| ventricular bigeminy | 82 |
| wolff parkinson white pattern | 79 |
| incomplete left bundle branch block | 77 |
| atrial flutter | 73 |
| anterior ischemia | 44 |
| ventricular hypertrophy | 29 |
| st elevation | 28 |
| supraventricular tachycardia | 27 |
| paroxysmal supraventricular tachycardia | 24 |
| ventricular trigeminy | 20 |
| complete heart block | 16 |
| 2nd degree av block | 14 |

### Datasets – Experiment B

In Experiment B, we used the Shandong Provincial Hospital (SPH) Database. The SPH database contains 25,770 12-lead ECG records with ECG signals ranging from 10 to 60 seconds long, together with accompanying diagnostic statements for each record \citealp{article1, sphdataset1, sphdataset2, sphdataset3, sphdataset4}. All ECG signals in the SPH database are sampled at 200Hz.

Experiment B consists of two sections. In the first section, we train from scratch the same 1D-CNN used as part of the CLIP-based model on the SPH dataset in order to obtain baseline results. The evaluation metrics used to obtain the baseline results were accuracy score, AUC-ROC, and AUC-PR. In the second section of Experiment B, the trained CLIP-based model from Experiment A was evaluated on the SPH dataset. The results and performance obtained by the CLIP-based model were then compared to the baseline results to evaluate whether a CLIP-based framework and model offers any benefits over simply just training a 1D-CNN image encoder.

## Data pre-processing

In Experiments A and B, the raw ECG signal for each ECG image-text pair was down sampled from its original frequency to 128 Hz. After down sampling the ECG signals, the ECG signals were then normalized by calculating the average mean and average standard deviation for all samples. The down sampled dataset is then filtered to exclude any diagnostic classes with fewer than 100 samples to ensure that sufficient data was available to train the models in both experiments.

## Model architecture

### Text Encoder – BioClinicalBERT

In Experiment A, the text encoder used as part of the CLIP-based model was BioClinicalBERT. BioClinicalBERT is a variant of the BioBERT base model which is additionally trained on all notes from MIMIC III, a database of electronic health records from ICU patients at Beth Israel Hospital in Boston, MA and which contains over 880 million words [31]. The pretrained BioClinicalBERT is publicly available through HuggingFace.

The implementation of the BioClinicalBERT text encoder in Python3 can be seen in the Figure below.

* Task: For LaTex, insert code correctly within a code block.

### Image Encoder – 1D-CNN or [SOME OTHER MODEL, TBD]

## Model training

In Experiment A, the CLIP-based model consisting of a 1D-CNN image encoder and the BioClinicalBERT text encoder was trained and validated on the PTB-XL dataset. The CLIP-based model was trained over 20 epochs with a learning rate of 0.001, Adam optimizer, a StepLR scheduler,

**Table 1:** Hyperparameters for Experiment A

|  |  |
| --- | --- |
| **Hyperparameter Type** | **Value** |
| Epochs | 20 |
| Optimizer | torch.optim.Adam |
| Projection Dimension | 64/128/256/512/768/1536 |
| Projection Head Learning Rate | 0.001 |
| Image Encoder Learning Rate | 0.001 |
| Image Encoder Type | 1D-CNN |
| Text Encoder Type | BioClinicalBERT |
| Temperature | 1.0 |
| ECG Signal Sample Rate (Hz) | 128 |
| Dropout | 0.0 |
| Criterion (Loss Function) | Contrastive Loss |

**Table 2:** Hyperparameters for Experiment B

|  |  |
| --- | --- |
| **Hyperparameter Type** | **Value** |
|  |  |

* Task: Create LaTex tables for Tables 1 and 2.

# Results

### 4.1 Results – Experiment A

The trained CLIP-based model demonstrated good performance during the training and validation phases. Over 20 epochs, the CLIP-based model achieved average AUC-ROC scores of 0.711 and 0.721 on the training and validationdata subsets. In addition, the average AUC-PR scores achieved were 0.073 and 0.073 on the training and validationdata subsets. The average accuracy score across all dataset labels in the training and validationdata subsets was 0.188 and 0.165, respectively. It was observed that class labels with a greater number of samples achieved higher scores across the evaluation metrics whereas class labels with fewer samples did not achieve high scores in evaluation metrics. A full summary of these results is summarized [here](https://github.com/naavie/Scientific-ECG-PUBLIC/blob/main/Experiment%20A/expA_history_256%20(1).csv).

When the trained CLIP-based model was evaluated on unseen classes of ECGs, it was observed that performance was poor suggesting that the model was struggling to generalize to unseen data.

### 4.2 Results – Experiment B

In Experiment B, the same image encoder from the CLIP-based model was trained on the Georgia dataset. A summary of the baseline results can be found here (hyperlink needed)

The trained CLIP-based model from Experiment B performed

# Discussion

# References

1. A large-scale multi-label 12-lead electrocardiogram database with standardized diagnostic statements. / H. Liu [et al.] // Sci Data. — 2022. — Vol. 9. — DOI: https://doi.org/10.1038/s41597-022-01403-5.

2. Ahsan M.M. Luna A.S. S. Z. Machine-Learning-Based Disease Diagnosis: A Comprehensive Review // Healthcare (Basel, Switzerland). — 2022. — Vol. 10, no. 3. — P. 1461–1471. — PMID: 35327018.

3. An Open Access Database for Evaluating the Algorithms of Electrocardiogram Rhythm and Morphology Abnormality Detection / F. Liu [et al.]. — 2018. — Journal of Medical Imaging and Health Informatics.

4. Artificial intelligence-enhanced electrocardiography in cardiovascular disease management / C. K. Siontis [et al.] // Nature Reviews Cardiology. — 2021. — Vol. 18. — P. 465–478. — DOI: 10.1038/s41569-020- 00503-2.

5. Automatic multilabel electrocardiogram diagnosis of heart rhythm or conduction abnormalities with deep learning: a cohort study / H. Zhu [et al.] // The Lancet. Digital health. — 2020. — Vol. 2, no. 7. — P. 348–357. — DOI: 10.1016/S2589-7500(20)30107-2.

6. Automatic Premature Ventricular Contraction Detection Using Deep Metric Learning and KNN / J. Yu [et al.] // Biosensors. — 2021. — Vol. 11, no. 3. — DOI: ttps://doi.org/10.3390/bios11030069.

7. Chhabra L. S. Y. Electrocardiogram // StatPearls [Internet]. — 2023.

8. Classification of 12-lead ECGs: the PhysioNet/Computing in Cardiology Challenge 2020 / P. Alday [et al.] // Institute of Physics and Engineering in Medicine. — 2020. — Vol. 41, no. 12. — DOI: <https://doi.org/10.1088/1361-6579/abc960>.

9. CLIP in Medical Imaging: A Comprehensive Survey / Z. Zhao [et al.] // arXiv. — 2023. — DOI: arXiv:2312.07353.

10. CLIP: Connecting text and images / A. Radford [et al.]. — 2021.

11. Cook A.D. Oh S.Y. P. V. Accuracy of Physicians’ Electrocardiogram Interpretations: A Systematic Review and Meta-analysis // JAMA internal medicine. — 2020. — Vol. 180, no. 11. — P. 1461–1471. — PMID: 32986084.

12. Diagnostic statement dictionary / H. Liu [et al.]. — 05/2022. — DOI: 10.6084/m9.figshare.17912507. v1. — URL: https://springernature.figshare.com/articles/dataset/diagnostic\_statement\_ dictionary/17912507 ; 12-lead electrocardiogram; ECG; Multi-label; ECG abnormalities. Figshare.

13. ECG records / H. Liu [et al.]. — 2022. — DOI: 10 . 6084 / m9 . figshare . 17912444 . v1. — URL: https://doi.org/10.6084/m9.figshare.17912444.v1. Figshare.

14. Electrocardiography: A Technologist’s Guide to Interpretation / C. Tso [et al.] // Journal of Nuclear Medicine Technology. — 2015. — Vol. 43, no. 4. — P. 247–252. — DOI: 10.2967/jnmt.115.163501.

15. ETP: LEARNING TRANSFERABLE ECG REPRESENTATIONS VIA ECG-TEXT PRE-TRAINING / C. Liu [et al.] // arXiv. — 2023. — DOI: arXiv:2309.07145v1.

16. Frozen Language Model Helps ECG Zero-Shot Learning / J. Li [et al.] // arXiv. — 2023. — DOI: arXiv:2303.12311.

17. Global, regional, and national age-sex specific mortality for 264 causes of death, 1980-2016: a systematic analysis for the Global Burden of Disease Study 2016 // THE LANCET. — 2017. — Vol. 390, 10100. — P. 1151–1210. — PMID: 28919116.

18. Heart Disease and Heart Attack. — Centers for Disease Control, Prevention, 2021. — URL: https: //www.cdc.gov/policy/polaris/healthtopics/heartdisease/index.html ; Accessed: 2024-01-15. [Electronic resource].

19. KecNet: A Light Neural Network for Arrhythmia Classification Based on Knowledge Reinforcement / P. Lu [et al.] // Journal of Healthcare Engineering. — 2021. — DOI: https://doi.org/10.1155/2021/ 6684954. — eprint: 6684954.

20. Learning Transferable Visual Models From Natural Language Supervision / A. Radford [et al.] // Proceedings of Machine Learning Research. — 2021. — Vol. 139. — P. 8748–8763.

21. Lui J.N. et al. Impact of New Cardiovascular Events on Quality of Life and Hospital Costs in People With Cardiovascular Disease in the United Kingdom and United States // Journal of the American Heart Association. — 2023. — Vol. 12, no. 19. — ISSN e030766. — DOI: 10.1161/JAHA.123.030766. — URL: <https://www.ahajournals.org/doi/10.1161/JAHA.123.030766>.

22. Machine learning-based detection of cardiovascular disease using ECG signals: performance vs. complexity / H. Pham [et al.] // Front. Cardiovasc. Med. — 2023. — Vol. 10. — DOI: 10.3389/fcvm.2023. 1229743. 20

23. Mamun K.R.M.M. E. T. AI-Enabled Electrocardiogram Analysis for Disease Diagnosis // Applied System Innovation. — 2023. — Vol. 6, no. 5. — DOI: <https://doi.org/10.3390/asi6050095>.

24. Mapping from Chinese ECG statements to AHA codes / H. Liu [et al.]. — 05/2022. — DOI: 10.6084/ m9.figshare.19738468.v1. — URL: https://springernature.figshare.com/articles/dataset/ Mapping\_from\_Chinese\_ECG\_statements\_to\_AHA\_codes/19738468 ; 12-lead electrocardiogram; ECG; Multi-label; ECG abnormalities. Figshare.

25. Mensah A.G. Roth A.G. F. V. The Global Burden of Cardiovascular Diseases and Risk Factors: 2020 and Beyond // Journal of the American College of Cardiology. — 2019. — Vol. 74, no. 20. — P. 2529–2532. — ISSN 0735-1097. — DOI: 10.1016/j.jacc.2019.10.009. — URL: https://www.sciencedirect.com/ science/article/pii/S0735109719379288.

26. OpenStax College Anatomy & Physiology //. — OpenStax College, 2013. — URL: https://commons. wikimedia.org/w/index.php?curid=30148227.

27. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals / A. Goldberger [et al.] // Circulation [Online]. — 2000. — Vol. 101, no. 23. — P. 215–220.

28. Pollock D., Makaryus N. Physiology, Cardiac Cycle // StatPearls. — 2022.

29. Prutkin M. ECG tutorial: Basic principles of ECG analysis // UpToDate. — 2023.

30. PTB-XL, a large publicly available electrocardiography dataset / P. Wagner [et al.]. — 2020. — URL: https://www.nature.com/articles/s41597-020-0495-6. Scientific Data.

31. Publicly Available Clinical BERT Embeddings / E. Alsentzer [et al.] // Proceedings of the 2nd Clinical Natural Language Processing Workshop. — Association for Computational Linguistics, 2019. — P. 72– 78. — DOI: 10.18653/v1/W19-1909. — URL: <https://www.aclweb.org/anthology/W19-1909>.

32. Scikit-learn: Machine Learning in Python / F. Pedregosa [et al.] // Journal of Machine Learning Research. — 2011. — Vol. 12. — P. 2825–2830.

33. State-of-the-Art Deep Learning Methods on Electrocardiogram Data: Systematic Review / G. Petmezas [et al.] // JMIR Medical Informatics. — 2022. — Vol. 10, no. 8. — PMID: 35327018.

34. The attributes of ECG records / H. Liu [et al.]. — 2022. — DOI: 10.6084/m9.figshare.17912441.v1. — URL: https://doi.org/10.6084/m9.figshare.17912441.v1. Figshare.

35. Will Two Do? Varying Dimensions in Electrocardiography: The PhysioNet/Computing in Cardiology Challenge 2021 / M. Reyna [et al.] // PhysioNet. — 2022. — DOI: <https://doi.org/10.13026/34va7q14>.

36. World Heart Report 2023. — World Heart Federation, 2023. — URL: https : / / world - heart - federation.org/resource/world-heart-report-2023/ ; Accessed: 2023-10-15. [Electronic resource].

37. Yanowitz G. 1. The Standard 12 Lead ECG // ECG LEARNING CENTER. —.

38. Zero-Shot Classification. — 2023. — URL: https : / / huggingface . co / tasks / zero - shot - classification

39. Romera-Paredes, B., Torr, P.H.S. (2017). An Embarrassingly Simple Approach to Zero-Shot Learning. In: Feris, R., Lampert, C., Parikh, D. (eds) Visual Attributes. Advances in Computer Vision and Pattern Recognition. Springer, Cham. https://doi.org/10.1007/978-3-319-50077-5\_2

40. Xian, Yongqin, et al. "Zero-shot learning—a comprehensive evaluation of the good, the bad and the ugly." *IEEE transactions on pattern analysis and machine intelligence* 41.9 (2018): 2251-2265.